Bitcoin price predication: linear regression

Prepared by:

Rawabi Almutairi

Fatimah Alqahtani

Raghad Alsaqer

Background

The Bitcoin scheme carries attributes of a payment system in that it facilitates the transfer of value between parties. Unlike traditional payment systems, which typically involve the transfer of value denominated in a sovereign currency such as the US dollar, Bitcoin has its own metric for value and is not a liability on any balance sheet. Having the power to move bitcoins inside the Bitcoin ecosystem is all that it takes to own bitcoins. As such, Bitcoin has no intrinsic value. Rather, a bitcoin's value is derived mainly from its use for making payments in the Bitcoin system, and from the purpose of accruing gains from bitcoins' possible appreciation. To our knowledge, Bitcoin has no legal tender status in any jurisdiction.11 Moreover, some economists have questioned whether bitcoins meet the standard attributes of money.

The year after February 2013 saw exceptional growth for Bitcoin. More than 64, 000 establishments throughout the world were reported to accept bitcoin payments as of October 7, 2014, and the exchange rate was more than US$300 to the bitcoin, which is more than 50 times greater than it was 24 months before. These examples show that Bitcoin's seemingly attractive potential may be accompanied with risks whose nature and magnitude are poorly known, if at all. Bitcoin, like all cryptocurrencies, is a complicated concept. Cryptography, distributed algorithms, and incentive-driven behaviour are used in its implementation. Our aims in this project:

We aims to introduce a predictive model for the price of Bitcoin in addition, explore a couple significant underlying features of the mode.

Question/Need

- What are the features we are interested in?
- We need a predictive model for the bitcoin price rate.
- What significant underlying features of the model?
- Which features have more impact with the target feature?

- Is there any multicollinearity? if there which features.

Model overview

The model is a multiple linear regression model, which means that it predicts a single dependent variable using more than one explanatory variable. There are three essential features (independent variables) that are highly correlated with Bitcoin's price (dependent variable).

Features Considered Include

The goal is to find attributes that have a strong association to Bitcoin but are located outside of the Bitcoin universe (an example would be cryptocurrency universe market capitalization, of which Bitcoin comprises 35% according to coinmarketcap.com).

The Following Features are taken into account:

Bitcoin Related

- Cryptocurrency universe market
- capitalization
- Ethereum price
- Volume
- Number of transactions
- Average block size
- Transaction fees
- Unique addresses
- Hash rate

Market related

- Price of Gold
- Nasdaq Composite Index

Others

- Google search interest

Tools

- Python (html5, matplotlib, Numpy, Pandas, sklearn ,seaborn ,etc…)
- CoinMarketCap website.
- Google search engine.

Framework

Dataset

1-DataFrame Construction

The data was obtained from various sources. First, the data for Bitcoin was pulled by Beautiful Soup: BTC, ETH. We also used quandl API to withdraw the rest of the features:

Price, volume,Txn fees ,cost per txn,num txns,txns per block,blk size,unique addys,hash rate,difficulty,NASDAQ Composite, Nasdaq GOLD Index, Google trend. In addition, the data came in the Time Period 2017-01-20 to 2021-10-29.

Index	BTCPrice	logBTCPrice	logETHPrice	logBTCVol	logTxFees	logCostperTxn	logNoTxns	logAvgBlkSz	logUniqueAddresses	logHashRate	logNasdaq	logGold	logInterest	Interest	TxFees	Nasdaq
2021-06-01 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-08 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-15 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-22 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-29 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816

Figure 1: Sample of DataFrame.

2- EDA

First, we dealt with the empty value by fillna(method='ffill') and delete unnecessary columns. Most of these features were also transformed onto a natural log scale as they relate exponentially / multiplicatively, so this transformation allows a more linear.

	logBTCPrice	logNasdaq	logInterest	logTxFees
logTxFees	0.527536	0.245675	0.707093	1.000000
logInterest	0.731891	0.352646	1.000000	0.707093
logNasdaq	0.779311	1.000000	0.352646	0.245675
logBTCPrice	1.000000	0.779311	0.731891	0.527536

Figure2: Multiplicatively for three features.

In addition, we did evaluate the correlation of the majority of features. We found the 3 core features for the model:

- Nasdaq.
- Interest.
- TxFees.

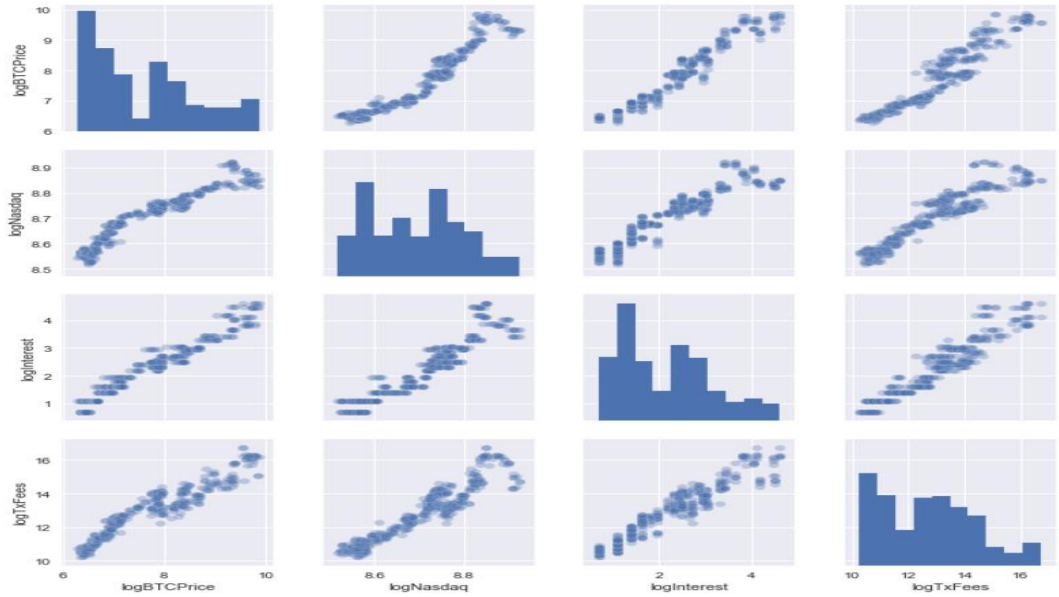

Figure3: Three core features.

Merge the Three Features to be inserted into the Model.

Index	ɔgBTCPric	ogNasdac	ogInteres	logTxFees
2021-06-01 00:00:00	8.907	8.964	3.829	14.32
2021-06-08 00:00:00	8.907	8.964	3.829	14.32
2021-06-15 00:00:00	8.907	8.964	3.829	14.32
2021-06-22 00:00:00	8.907	8.964	3.829	14.32
2021-06-29 00:00:00	8.907	8.964	3.829	14.32

Figure4: The target and features.

3-Linear Regression Analysis

The dataset was divided as follows: Training dataset was given approximately 70% and Test dataset was 30%.

OLS Regression Results

Dep. Variable:	logBTCPrice	R-squared:	0.846
Model:	OLS	Adj. R-squared:	0.845
Method:	Least Squares	F-statistic:	1093.
Date:	Fri, 22 Oct 2021	Prob (F-statistic):	9.14e-242
Time:	09:56:06	Log-Likelihood:	-86.459
No. Observations:	600	AIC:	180.9
Df Residuals:	596	BIC:	198.5
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-28.6365	1.013	-28.280	0.000	-30.625	-26.648
logInterest	0.5346	0.025	21.446	0.000	0.486	0.584
logNasdaq	3.9967	0.115	34.680	0.000	3.770	4.223
logTxFees	0.0146	0.014	1.075	0.283	-0.012	0.041

Omnibus:	8.798	Durbin-Watson:	0.080
Prob(Omnibus):	0.012	Jarque-Bera (JB):	8.779
Skew:	-0.269	Prob(JB):	0.0124
Kurtosis:	3.247	Cond. No.	1.46e+03

Figure5: Model summary.

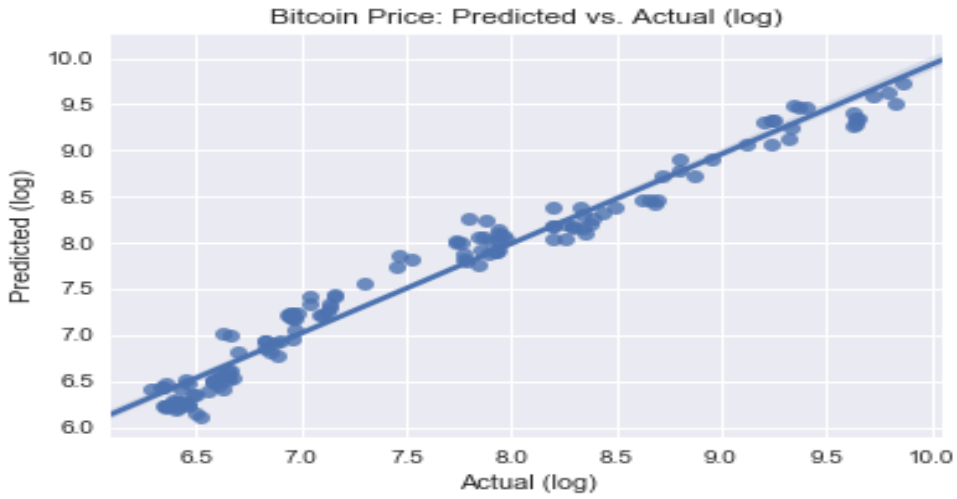

Figure 6: Bitcoin price linear regression.

Training dataset and find this result

- Trian (x,y)=0.8399367307539943
- Test (x,y)=0.8568224150791021

Regularization

The model is correlated with the price of Bitcoin, achieved with only 3 features. Even with a further reduction of features the model remains correlated. Per the simply RidgeCV analysis below, the model is also not overfit.

- Trian (x,y)=0.8397597320509806
- Test (x,y)= 0.8568917069026165